Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Marcus CL, Moore RH, Rosen CL, et al. A randomized trial of adenotonsillectomy for childhood sleep apnea. N Engl J Med 2013;368:2366-76. DOI: 10.1056/NEJMoa1215881

SUPPLEMENTARY APPENDIX

SUPPLEMENT TO: A RANDOMIZED CONTROLLED TRIAL OF SURGERY FOR CHILDHOOD SLEEP APNEA

Carole L. Marcus, MBBCh[1]; Reneé H. Moore, PhD[2,3], Carol L. Rosen, MD[4], Bruno Giordani, PhD[5,6], Susan L. Garetz, MD[7] , H. Gerry Taylor, Ph.D[4], Ron B. Mitchell, MD[8,9], Raouf Amin, MD[10], Eliot S. Katz,MD[11], Raanan Arens, MD[12], Shalini Paruthi, MD[13], Hiren Muzumdar, MD[12], David Gozal, MD[14], Nina Hattiangadi Thomas, PhD[15], Janice Ware, PhD[16], Dean Beebe, PhD[10], Karen Snyder, MS[13], Lisa Elden, MD[17], Robert C. Sprecher, MD[18], Paul Willging, MD[19], Dwight Jones, MD[20], John P. Bent, MD[21], Timothy Hoban, MD[6], Ronald D. Chervin, MD, MS[6],Susan S. Ellenberg, PhD[2], Susan Redline, MD,MPH[22]

[1] Department of Pediatrics, Sleep Center, Children's Hospital of Philadelphia; University of Pennsylvania, Philadelphia, PA
[2] Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA
[3] Department of Statistics, North Carolina State University, Raleigh, NC
[4] Department of Pediatrics, Rainbow Babies & Children's Hospital, University Hospitals Case Medical Center, Case Western Reserve University, Cleveland, OH
[5] Department of Psychiatry and Psychology, University of Michigan, Ann Arbor, MI
[6] Departments of Pediatrics and Neurology, University of Michigan, Ann Arbor, MI
[7]Department of Otolaryngology, University of Michigan, Ann Arbor, MI
[8] Department of Otolaryngology, Cardinal Glennon Children's Medical Center, St Louis, MO
[9]Departments of Otolaryngology and Pediatrics, UT Southwestern Medical Center, Dallas, TX
[10] Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH
[11] Department of Pediatrics, Boston Children's Hospital, Harvard Medical School, Boston, MA
[12]Department of Pediatrics, Children's Hospital at Montefiore and Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, NY
[13] Department of Pediatrics, Cardinal Glennon Children's Medical Center, Saint Louis University, St Louis, MO
[14] Department of Pediatrics, University of Chicago, Chicago, IL
[15] Department of Child and Adolescent Psychiatry and Behavioral Sciences, Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA
[16] Departments of Medicine and Psychiatry, Boston Children's Hospital, Harvard Medical School, Boston, MA
[17] Department of Otolaryngology, Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA
[18] Department of Otolaryngology, Rainbow Babies & Children's Hospital, University Hospitals Case Medical Center, Case Western Reserve University, Cleveland, OH
[19] Department of Otolaryngology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH
[20] Department of Otorhinolaryngology, Boston Children's Hospital, Harvard Medical School, Boston, MA
[21] Department of Otolaryngology, Head & Neck Surgery, Children's Hospital at Montefiore and Montefiore Medical Center, Albert Einstein College of Medicine, Bronx, NY
[22] Department of Medicine, Brigham and Women's Hospital and Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA

For the **Ch**ildhood **A**denotonsillectomy **T**rial (CHAT): Boston Children's Hospital, Harvard University, Boston MA (Eliot Katz, MD; Janice Ware, PhD; Dwight Jones, MD); Brigham and Women's Hospital and Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA (Susan Redline, MD MPH; Rui Wang PhD); Cardinal Glennon Children's Hospital, Saint Louis University, St. Louis, MO (Ron Mitchell, MD; Shalini Paruthi, MD; Karen Snyder, MS); University of Pennsylvania/Children's Hospital of Philadelphia, (Carole Marcus, MBBCh; Nina H. Thomas, PhD; Lisa Elden, MD); Cincinnati Children's Medical Center, University of Cincinnati, Cincinnati, OH (Raouf Amin, MD; Dean Beebe, PhD; Paul Willging, MD); Montefiore Children's Hospital, Albert Einstein College of Medicine, Yeshiva University, Bronx, NY (Raanan Arens, MD; Hiren Muzumdar, MD; Shelby Harris, PsyD CBSM); Rainbow Babies and Children's Hospital, University Hospitals Case Medical Center, Case Western Reserve University School of Medicine, Cleveland, OH (Carol Rosen, MD; H. Gerry Taylor, PhD; Robert Sprecher, MD; James Arnold, MD); University of Kentucky, Louisville, Kentucky (David Gozal, MD); University of Michigan, Ann Arbor, MI (Ronald Chervin, MD; Susan Garetz, MD, Bruno Giordani, PhD; Tim Hoban, M.D.); University of Pennsylvania, Philadelphia, PA (Susan Ellenberg PhD; Reneé H. Moore, PhD; Kim Lacy, RN, BSN)

TABLE OF CONTENTS FOR WEB-BASED SUPPLEMENTARY APPENDIX

METHODS: Study Design and Participating Sites……page 5

METHODS: Quality Control……..……...page 6

METHODS: Study Sample and Recruitment……..…….…page 6

METHODS: Subject characteristics…….page 7

METHODS: Polysomnography…….……..…page 7

METHODS: Cognitive, behavioral and symptom evaluations…..……..…page 8

METHODS: Adenotonsillectomy……..…page 8

METHODS: Other Outcomes……..……page 8

METHODS: Safety monitoring……..…….page 9

METHODS: Statistical considerations……...…….page 10

RESULTS: Participant Flow…….…….……page 11

RESULTS: Intervention effects among those treated per protocol…….page 12

RESULTS: Subgroup differences between arms…….....page 12

RESULTS: Other exploratory analyses……..page 13

RESULTS: Adverse events and treatment failures…….page 13

RESULTS: Otolaryngology Follow-up In the WWSC group……..page14

DISCUSSION: Safety and treatment failures.……...…page 14

DISCUSSION: Ethical considerations…….…….……..…page 14

Figure S1 Change in ln AHI by baseline AHI……..…….page 16

Table S1: Baseline Characteristics of entire sample and completed sample

……...……..pages 17-19

Table S2. Outcome measures for the early adenotonsillectomy (eAT) compared to the Watchful Waiting With Supportive Care (WWSC) groups……...page 20-24

Table S3 Adverse Events……..pages 27-28

REFERENCES……..…..…..……page 26-27

SUPPLEMENTARY METHODS

Study Design and Participating Sites

This was a multicenter, single-blind, randomized controlled trial that recruited children with symptoms of OSAS from primary care, otolaryngology, and sleep clinics at 7 academic sleep centers (Children's Hospital of Philadelphia, Philadelphia, PA; Cincinnati Children's Medical Center, Cincinnati, OH; Kosair Children's Hospital (KCH), Louisville, KY; Rainbow Babies and Children's Hospital, Cleveland, OH; Children's Hospital, Boston, MA; Cardinal Glennon Children's Hospital, St. Louis, MO; and Montefiore Medical Center, Bronx, NY). One site (KCH) was removed from the study after its principal investigator relocated. Children underwent polysomnography, cognitive and behavioral testing and other clinical and laboratory evaluations to evaluate cognition, behavior, quality of life, sleep parameters, blood pressure, growth, and metabolic profile at baseline and 7 months; this paper addresses the first four domains.

At the baseline visit, children were evaluated for co-morbid conditions that could exacerbate OSAS, such as allergies and poorly controlled asthma, and were referred for appropriate treatment as needed according to local clinical practice; these referrals did not affect randomization. All children/caregivers received information on sleep hygiene using standardized educational materials (including brochures, fun pads, etc.) that identified the need for regular sleep routines, age-appropriate sleep duration, avoidance of caffeine prior to bedtime, etc., and were provided information on healthy lifestyle (nutrition, exercise). Children also were provided with saline nasal spray to be used nightly. Children were randomized to either early adenotonsillectomy (eAT; surgery within 4 weeks of randomization) or to Watchful Waiting with Supportive Care (WWSC). Repeat polysomnography and cognitive and behavioral testing were performed at approximately 7 months following randomization. Children randomized to WWSC were scheduled for reevaluation for surgery by an otolaryngologist after the 7 month observation period, and received treatment as clinically indicated. The primary outcome was change in the

Attention/Executive Functioning Domain score from the Developmental NEuroPSYchological Assessment (NEPSY A/E.[1] This test has well-established psychometric properties with a high stability coefficient.[1]

Quality Control

To reduce bias, personnel involved in conducting psychometric evaluations and measuring outcomes, as well as study investigators (other than the surgeons) were blinded to randomization assignments. The study was single-blind, i.e., investigators were blinded but concerns about the feasibility of sham surgery precluded blinding of subjects' families. Study quality control procedures included central training of site coordinators and investigators; certification of research personnel for all specialized testing procedures and for data entry; and site visits. Data were reviewed on a regular basis by surgical, neuropsychological and polysomnography quality control subcommittees, as well as by the Data Coordinating Center (University of Pennsylvania). The study was approved by the Institutional Review Board of each institution. Written permission (informed consent) was obtained from caregivers, and assent from children $\geq$ 7 years of age.

Study sample and recruitment

Children referred for evaluation of OSAS, tonsillar hypertrophy or frequent snoring were recruited from pediatric sleep centers, otolaryngology and other pediatric clinics, and community sources. Children were eligible for study entry if they were 5-9.9 years of age, had a history of snoring, had tonsillar hypertrophy $\geq$ 1 on a scale of 0-4,[2] and were considered to be candidates for AT by an otolaryngologist. Exclusion criteria included a history of recurrent tonsillitis, extreme obesity (body mass index [BMI] z-score $\geq$ 3), therapy for failure to thrive, medications for psychiatric or behavioral disorders (including attention deficit hyperactivity disorder), and known medical conditions likely to affect the airway, cognition or behavior. Children were

screened further by standardized polysomnography and cognitive testing. Children with OSAS without prolonged desaturation, defined as an obstructive apnea hypopnea index (AHI) between 2-30/hr or an obstructive apnea index OAI) between 1-20/hr, and time with arterial oxygen saturation (SpO_2) < 90% that was < 2% of total sleep time, and those without evidence of marked intellectual impairment as assessed by a General Conceptual Ability (GCA) score on the Differential Ability Scales (DAS II)[3] > 55 at the baseline examination, were eligible for randomization. The AHI and OAI cutoffs were based on published normative pediatric data.[4-8] Classification of race/ethnicity was through caregiver report, using the NIH classification system.

Subject characteristics

Children's nutritional status was characterized as failure to thrive (weight or height < 5th percentile),[9] normal weight, overweight (BMI 85-94th percentile)[10] and obese (BMI $\geq$95th percentile) [10]. Family income was categorized as < or $\geq$ $30,000 per year or not reported.

Polysomnography

Children underwent full, in-laboratory polysomnography by study-certified technicians according to a standardized protocol, using similar sensors, and following American Academy of Sleep Medicine (AASM) guidelines.[11] Scoring was performed according to the AASM pediatric criteria, by certified technologists blinded to all other study data at a central Polysomnography Reading Center (Case Western Reserve University/Brigham and Women's Hospital). The AHI was defined as the sum of all obstructive and mixed apneas, plus hypopneas associated with a 50% reduction in airflow and either a $\geq$ 3% desaturation or electroencephalographic arousal, divided by hours of total sleep time; the OAI as all obstructive apneas per sleep hour; and the Oxygen Desaturation Index (ODI) as the number of oxyhemoglobin desaturations $\geq$3% per sleep hour. The inter-scorer reliability, assessed by the intra-class correlation coefficient, was 0.98 for the AHI, and for other polysomnography variables ranged from 0.75 to 0.99.

Cognitive, behavioral and symptom evaluations

Neuropsychological testing was performed during a morning visit, on a separate day from the polysomnogram to avoid the influence of atypical sleep related to overnight monitoring. Tests were administered by psychometrists blinded to the polysomnographic results. A centrally trained, licensed psychologist supervised each psychometrist. At the baseline and follow-up examinations, caregivers were asked to complete standardized survey instruments. Children were randomized to receive one of two alternative test batteries, which presented the same tests but in a differing order. Within children, test order was maintained for baseline and follow-up assessments.

In association with each baseline and follow-up visit, the child's teacher was mailed behavioral assessments to complete.

Adenotonsillectomy

Participating otolaryngologists viewed a training video. Complete bilateral tonsillectomy and removal of obstructing adenoid tissue was performed by standard surgical techniques including cold dissection, monopolar electrocautery, coblation or micro-debrider, with variation according to surgeon preference and not by patient characteristics. Intra-operative photographs were obtained on every 10th subject at each site and were reviewed for adequacy of lymphoid tissue removal by the surgical quality control committee chair (SG).

Other Outcomes

Other outcomes reported here include polysomnographic indices and changes in summary or composite scores measuring behavior, symptoms and quality of life:

1. Behavior, by the parent rating on the *Conners' Parent Rating Scale-Revised: Long version Global Index* (CGI T score), a two-factor score comprising the Restless-Impulsive and Emotional-Lability factor sets[12], and by the *Behavior Rating Inventory of Executive Function (BRIEF) Global Executive Composite (GEC)* T score[13], comprising summary measures of behavioral regulation and metacognition. Teacher ratings from parallel instruments (the CGI short version and BRIEF Teacher Report Form) were also evaluated.
2. Symptoms of OSAS, by the total score of the *Pediatric Sleep Questionnaire (PSQ) Sleep Related Breathing Disorder Scale (SRBD)*.[14]
3. Sleepiness, the *Epworth Sleepiness Scale* modified for children.[15]
4. Global quality of life, by the parental total score from the *Pediatric Quality of Life Inventory* (PedsQL)[16], and disease-specific quality of life, assessed by the total score of the *OSA-18*[17], a composite of OSAS-related symptoms and quality of life.
5. The Differential Ability Scales II (DAS), a measure of generalized intellectual functioning.

Safety monitoring

Families were contacted by telephone every 2 months to elicit reports of adverse events, and an interim visit was conducted at month 3. A central, independent, unblinded medical monitor (Timothy Hoban, M.D.) reviewed all serious and unexpected adverse events, as well as Treatment Failures which were defined as serious changes in clinical status potentially related to inadequately treated OSAS that might require additional/alternative therapies. Research coordinators identified potential Treatment Failures through regular contact with the participants' families, which included assessment of changes in the child's clinical status. Once identified, relevant clinical information was summarized and provided to the Data Coordinating Center Project Manager and to the Medical Monitor who adjudicated these events. Examples of potential treatment failures included: new academic or behavioral problems resulting in a

recommendation for grade retention; recurrent bacterial tonsillitis; new clinical diagnosis of cor pulmonale; and development of failure to thrive. A series of case studies of potential Treatment Failures was prepared and used to train the Medical Monitor, study investigators and staff on consistent identification of Treatment Failures. An independent Data and Safety Monitoring Board (DSMB) reviewed interim data on safety and study quality; members included: Lynn Taussig, M.D. (Chair); Thomas Anders, M.D.; Julie Buring, Sc.D.; Karina Davidson, Ph.D.; Estelle Gauda, M.D.; Steven Piantadosi, M.D., Ph.D.; Bennett Shaywitz, M.D.; Benjamin Wilfond, M.D.; Tucker Woodson, M.D.; and Robert Zeiger, M.D.

Statistical considerations

We determined that a sample size of 400 children, randomized 1:1 between eAT and WWSC, would permit detection of an effect size for the primary endpoint of the NEPSY A/E domain score of 0.32 (an effect size estimated from one prior study[18]) or greater with power of 90%. We planned to enroll a total of 460 children to compensate for dropouts. Randomization was performed centrally using a web-based system that required confirmation of eligibility criteria prior to providing the treatment assignment. Primary analyses included all randomized subjects for whom change in the primary outcome could be assessed, with the exception of 11 children from the site that terminated participation early and whose subjects were excluded from the primary analyses. Following the intention to treat principle, children who remained in the study but crossed over to the alternative treatment were included in their assigned treatment groups for primary analysis. The primary analysis was specified as an analysis of covariance adjusting for the stratification factors of age (5-7 versus 8-9 years old), race (African American versus other), weight status (overweight/obese versus non-overweight) and study site. Secondary outcomes were analyzed similarly. Additional analyses of primary and secondary outcomes were performed with adjustments for other factors expected to be prognostic, including baseline values of the outcome and AHI, family income and season. Variables with highly skewed

distributions were log transformed for analysis. Interaction tests were planned to assess whether race, obesity or baseline AHI influenced the effect of the intervention. A number of exploratory analyses also were conducted, including assessment of changes only in those who (i) received the assigned treatment; (ii) those with baseline NEPSY A/E scores in the lowest quartile; and (iii) those who experienced resolution of OSAS by polysomnography. We also tested for interactions between treatment arm and other potentially predictive covariates (i.e., site, income level). Sensitivity analyses using multiple imputation were used to assess the impact of the missing values for the primary outcome. For this analysis, an ANCOVA model including all randomized children except the 11 from the site excluded for administrative reasons was fit utilizing the SAS MI and MIANALYZE procedures. The MI procedure assumed a multivariate normal distribution and utilized a Markov Chain Monte Carlo (MCMC) method with a single chain to create 10 imputations (complete data sets), and computed the posterior mode with a noninformative prior via the EM algorithm. Two interim analyses were performed after 25% and 50% of subjects had completed their 7-month evaluation, and were reviewed by the DSMB.

SUPPLEMENTARY RESULTS

Study Flow

In total, 1447 children were identified as potentially eligible for this study. Of these, 1244 (86%) underwent screening polysomnography, which identified 594 (48%) children as meeting the study OSAS eligibility criteria; 49% of children were excluded as polysomnography findings were too mild, and 3% were excluded as polysomnography findings were too severe. No child required exclusion secondary to a low DAS II score. Of the eligible subjects, 464 children were randomized between January 2008 and September 2011 to either eAT (n=232) or WWSC (n=232), with the remaining children not being randomized due to other eligibility factors or changes in caregiver interest. Analyses excluded 11 children recruited from one site (KCH)

whose principal investigator relocated early in the study, and where follow-up data were available in too few subjects (N = 3) for meaningful analysis. Of the remaining 453 children, 35 children were lost to follow-up and 18 withdrew from the study, resulting in 7-month follow-up visits for 400 (88%) children, with 397 evaluable NEPSY A/E measurements. A comparison of children who completed the study with those who did not complete showed no significant differences in OSAS severity or baseline NEPSY A/E scores; significantly more African American children than others did not complete the study (p=0.04), but this trend was evident in both treatment arms.

Intervention effects among those treated per protocol

Due to caregiver decisions or because of recommendations by the external medical monitor, of the children completing the study, 16 children assigned to the WWSC group underwent AT prior to 7 months post-randomization, and 8 children randomized to eAT did not receive AT. Excluding these 24 children from analysis did not yield any appreciable changes in the study results.

Subgroup differences between arms

Models evaluating possible effect modification of treatment by race, obesity, median AHI and age were tested by including terms for interactions between treatment arm and each of these factors on each of the study outcomes. Neither obesity nor age group significantly modified treatment responses for any of the outcomes reported here. The relative improvements associated with eAT compared to WWSC were significantly lower for African American children compared to children of other ethnic/racial backgrounds for the caregiver completed CGI (-1.06 $\pm$ 10.85 vs. -0.98 $\pm$ 9.53 for eAT vs. WWSC in African Americans, and -4.84 $\pm$ 9.49 vs. 0.61 $\pm$ 9.22 for other racial groups; interaction between race and treatment p<0.01); BRIEF GEC T-score (-1.82 $\pm$ 8.86 vs. -0.30 $\pm$ 9.27 for African Americans, and -4.98 $\pm$ 7.69 vs. 1.17 $\pm$ 8.29 for

others; p<0.05); and PSQ-SRBD scale (-0.24 ± 0.19 vs. -0.04 ± 0.19 for African Americans, and -0.32 ± 0.16 vs. -0.02 ± 0.18 for others, p<0.01). Children with more severe OSAS at baseline (i.e., those with an AHI greater than the median AHI of 4.7/hr) who were assigned to eAT experienced a greater improvement in polysomnographic indices (figure 2), including AHI (p<0.001), ODI (p<.01), arousal index (p<0.05) and percentage of sleep time with CO_2 > 50 mm Hg (p<0.05). There was no interaction between OSAS severity (defined by median AHI) and treatment arm for the NESPY A/E or for the behavior or symptom outcomes. To further explore differential responses among those with severe OSAS, we tested for interaction between OSAS severity defined as a baseline AHI in the top quartile (AHI > 8.7/hr) and intervention group for effect on NEPSY A/E. We found no suggestion of a difference in NEPSY A/E change in children in the top quartile of AHI in the eAT compared to the WWSC group (NEPSY A/E change: eAT: 6.96 (SD 16.4; n=49) vs WWSC: 6.33 (SD 12.6; n=54) in analyses adjusted for stratification variable (p=0.81) or in models adjusting for the other covariates (p=0.48).

Other exploratory analyses

No consistent relationships were observed between changes in AHI with changes in study outcomes, indicating that the degree of physiologic improvement in OSAS as measured by this metric was not substantially predictive of changes in cognition or behavior. Analyses that additionally adjusted for season, obesity, income, sex, baseline AHI, and the baseline value of each outcome did not result in substantive differences as compared to the primary analyses (data not shown). Further, no appreciable differences in findings were observed for analyses restricted to children with baseline NEPSY A/E scores in the lowest quartile.

Adverse events and treatment failures

Of the 15 post-randomization serious adverse events, 6 occurred in children randomized to eAT, while 9 occurred in children in the WWSC group. Eight of the events were associated with

peri-operative complications (bleeding, dehydration, pain), 3 of which occurred in children randomized to WWSC but who “crossed-over” to surgery. The medical monitor adjudicated 9 treatment failures (also considered adverse events), all in the WWSC group, resulting in recommendations for early surgery. Treatment failures were attributed to: increased problems with sleep quality or sleepiness (n=3), school behavioral problems (n=1), morning headaches (n=1), asthma exacerbation (n=1), hypertension (n=1) and bacterial infections (n=2).

Otolaryngology Follow-up In the WWSC group

In the WWSC, data were available for 147 children evaluated by otolaryngology after their 7 month visit. Of these, 71% were considered by the otolaryngologist to be candidates for adenotonsillectomy at that time.

SUPPLEMENTARY DISCUSSION

Safety and treatment failures

Treatment with either eAT or WWSC could be associated with safety concerns. Surgery was associated with a 3% rate of serious adverse peri-operative complications (defined as requiring an additional operative procedure, hospitalization or prolonging hospitalization), a rate consistent with the results of a meta-analysis of post-operative bleeding[19], but was not associated with death or persistent disability in any child. A 3% rate of adjudicated treatment failures also was observed, which included events such as new behavioral problems, sleepiness, or recurrent pharyngitis; these only occurred in children randomized to WWSC. Thus, this trial supports the overall safety of both eAT and WWSC, but suggests the need for clinical monitoring of children initially managed with conservative medical management.

Ethical considerations

This study raised several ethical issues. One intervention exposed the child to general anesthesia and surgery with a known, albeit small, peri-operative morbidity and mortality,

whereas the alternative intervention could be perceived as withholding a clinically accepted standard therapy. The study was conducted because there is uncertainty regarding the utility of AT in treating childhood OSAS unaccompanied by prolonged desaturation. Although AT has been considered a standard intervention for childhood OSAS, there had never been a large, randomized controlled trial evaluating its efficacy. Changes in the clinical spectrum of patients referred for surgery, with an increasing prevalence of obese children[20, 21], necessitated evaluation of current clinical practice. A previous randomized controlled trial of AT for treatment of recurrent infection showed that generally held assumptions about the effectiveness of AT for treatment of infections were not borne out when studied under rigorous conditions.[22] It is only by performing large, controlled clinical trials that superior treatment strategies will be identified, and the management of health conditions in children will be improved.

Figure S1. The change in natural log apnea hypopnea index (AHI), stratified for baseline median AHI (4.7), is shown for the two study groups.

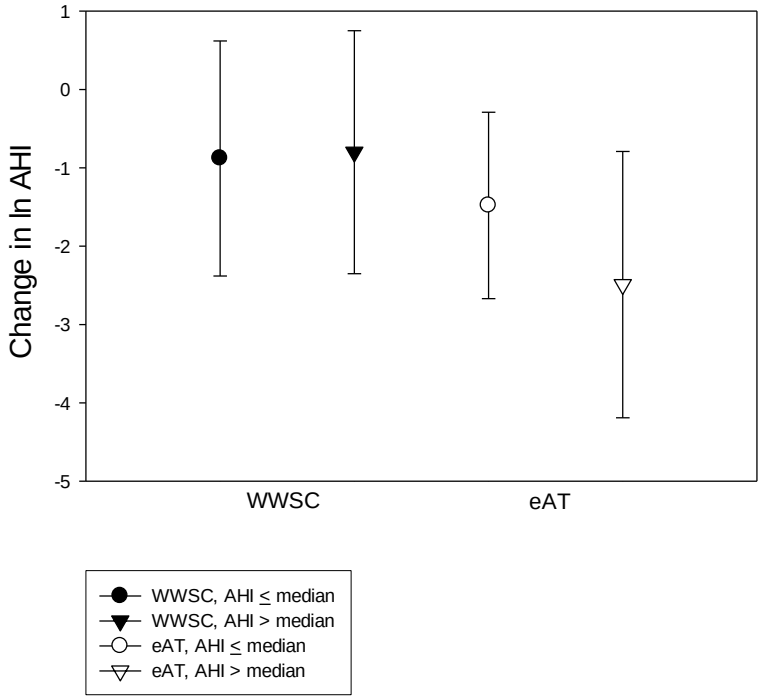

Table S1: Baseline characteristics of subjects

Characteristic	Randomized WWSC Arm	Randomized eAT Arm	Completed WWSC	Completed eAT
N	227	226	203	194
Age (yr)	6.5 ± 1.4	6.6 ± 1.4	6.5 ± 1.4	6.5 ± 1.4
Males	118 (52.0)	101 (44.7)	106 (52.2)	89 (45.9)
Race				
African American	123 (54.2)	126 (55.8)	108 (53.2)	103 (53.1)
Caucasian	81 (35.7)	75 (33.2)	76 (37.4)	67 (34.5)
Other	23 (10%)	26 (11.5%)	19 (9.4)	24 (12.4)
Hispanic Ethnicity	21 (9.3)	16 (7.1)	17 (8.4)	15 (7.7)
Anthropometric Measures				
Height (cm) Height z-score	124.8 ± 10.8 0.6 ± 1.0	125.5 ± 11.3 0.7 ± 1.0	124.7 ± 10.5 0.6 ± 1.0	125.1 ± 11.2 0.7 ± 1.0

Characteristic	Randomized WWSC Arm	Randomized eAT Arm	Completed WWSC	Completed eAT
Weight (kg)	30.4 ± 12.4	31.2 ± 13.0	30.1 ± 11.7	31.2 ± 13.1
Weight z-score	1.0 ± 1.2	1.0 ± 1.3	1.0 ± 1.2	1.0 ± 1.3
Weight class:				
Overweight or obese (BMI ≥ 85th percentile)	106 (46.7)	107 (47.3)	94 (46.3)	93 (47.9)
Obese (BMI ≥95th percentile)	76 (33.5)	74 (32.7)	67 (33.0)	68 (35.1)
Failure to thrive (weight <5th percentile)	4 (1.8)	6 (2.6)	3 (1.5)	4 (2.1)
Maternal Education < high school	69 (30.7)	71 (32.0)	64 (31.7)	62 (32.6)
Household Income: < $30,000/yr	135 (59.5)	132 (58.4)	82 (40.4)	73 (37.6)
Site				
Philadelphia	72 (31.7)	75 (33.2)	66 (32.5)	68 (35.0)
Cincinnati	40 (17.6)	39 (17.3)	30 (14.8)	28 (14.4)
Cleveland	60 (26.4)	64 (28.3)	57 (28.1)	59 (30.4)

Characteristic	Randomized WWSC Arm	Randomized eAT Arm	Completed WWSC	Completed eAT
St. Louis	30 (13.2)	30 (13.3)	30 (14.8)	25 (12.9)
New York	9 (4.0)	7 (3.1)	6 (3.0)	5 (2.6)
Boston	16 (7.0)	11 (4.9)	14 (6.9)	9 (4.6)

Data presented as mean ± SD or N (%). WWSC, Watchful Waiting With Supportive Care Group; eAT, early adenotonsillectomy group; BMI, body mass index. There are no statistically significant imbalances between eAT and WWSC.

Table S2: Outcome measures for the early adenotonsillectomy (eAT) compared to the Watchful Waiting With Supportive Care (WWSC) groups

		WWSC			eAT			
Outcome	**Parameter**	**Baseline**	**Month 7**	**Change from baseline to month 7**	**Baseline**	**Month 7**	**Change from baseline to month 7**	**p value**
PRIMARY OUTCOME								
NEPSY A/E	N	203	203	203	194	194	194	
	Mean (SD)	101.1 (14.6)	106.2 (15.0)	5.1 (13.4)	101.5 (15.9)	108.6 (15.5)	7.1 (13.9)	0.16
	Median (Q1,Q3)	102 (92,113)	107 (98,119)	5 (-3,14)	102 (92,110)	110 (98,121)	7.5 (-3,16)	
SECONDARY OUTCOMES								
Caregiver-reported Behavior and Executive Function								
Conners' CGI (caregiver)	N	199	199	199	193	193	193	
	Mean (SD)	52.6 (11.7)	52.4 (10.5)	-0.2 (9.4)	52.5 (11.6)	49.6 (10.8)	-2.9 (9.9)	0.01
	Median (Q1,Q3)	50 (43, 59)	50 (44, 58)	0 (-6, 6)	49 (44, 58)	47 (42, 54)	-2 (-7, 1)	
BRIEF GEC (caregiver)	N	197	197	197	195	195	195	
	Mean (SD)	50.1 (11.5)	50.5 (11.9)	0.4 (8.8)	50.1 (11.2)	46.8 (11.6)	-3.3 (8.5)	<0.001
	Median (Q1,Q3)	48 (42, 57)	48 (41, 60)	0 (-5, 6)	48 (42, 56)	43 (38, 55)	-2.0 (-8,1)	
Teacher-reported Behavior and Executive Function								

		WWSC			eAT			
Outcome	**Parameter**	**Baseline**	**Month 7**	**Change from baseline to month 7**	**Baseline**	**Month 7**	**Change from baseline to month 7**	**p value**
Conners' CGI (teacher)	N	109	109	109	103	103	103	
	Mean (SD)	55.1 (12.8)	53.7 (12.2)	-1.5 (10.7)	56.4 (14.4)	51.6 (12.0)	-4.9 (12.9)	0.04
	Median (Q1,Q3)	50 (46, 60)	50 (44, 58)	-2 (-5, 2)	52 (44, 66)	46 (44, 55)	-2 (-12,1)	
BRIEF GEC (teacher)	N	103	103	103	104	104	104	
	Mean (SD)	56.4 (11.7)	55.4 (13.5)	-1.0 (11.2)	57.2 (14.1)	54.2 (13.6)	-3.1 (12.6)	0.17
	Median (Q1,Q3)	55 (47, 63)	54 (45, 62)	0 (-6,5)	55.5 (45, 67)	51.5 (44, 62)	-1.5 (-12, 6)	
Symptoms & Health Quality of Life								
Modified Epworth Sleepiness Scale	N	202	202	202	196	196	196	
	Mean (SD)	7.4 (5.1)	7.1 (5.1)	-0.3 (4.1)	7.1 (4.6)	5.1 (4.4)	-2.0 (4.2)	<0.001
	Median (Q1,Q3)	7 (3,11)	6 (3,11)	0 (-2, 2)	6 (4,10)	4 (2, 7)	-2 (-4, 0)	
Pediatric Sleep Questionnaire – SRBD Scale	N	202	202	202	194	194	194	
	Mean (SD)	0.5 (0.2)	0.5 (0.2)	-0.0 (0.2)	0.5 (0.2)	0.2 (0.2)	-0.3 (0.2)	<.001
	Median (Q1,Q3)	0.5 (0.4, 0.6)	0.5 (0.3, 0.6)	-0.0 (-0.1, 0.1)	0.5 (0.3, 0.6)	0.2 (0.1, 0.3)	-0.3 (-0.4, -0.2)	
PedsQL (total score)	N	204	204	204	195	195	195	
	Mean (SD)	76.5 (15.7)	77.4 (14.9)	0.9 (13.3)	77.3 (15.3)	83.3 (15.1)	5.9 (13.6)	<0.001

		WWSC			eAT			
Outcome	**Parameter**	**Baseline**	**Month 7**	**Change from baseline to month 7**	**Baseline**	**Month 7**	**Change from baseline to month 7**	**p value**
	Median (Q1,Q3)	79 (66, 90)	79.7 (68, 89)	0.0 (-6, 9)	81 (66, 90)	88.0 (74.8, 95.7)	4.1 (-1.1,13.0)	
OSAS-18 (total score)	N	202	202	202	193	193	193	
	Mean (SD)	54.1 (19.2)	49.5 (20.3)	-4.5 (19.3)	53.2 (17.7)	31.8 (14.9)	-21.4 (16.5)	<.001
	Median (Q1,Q3)	52 (41, 66)	48 (34, 63)	-3.5 (-16, 9)	50 (40, 64)	27 (22, 37)	-21 (-31, -10)	
Polysomnography								
AHI (N/hr)	N	208	208	208	199	199	199	
	Mean (SD)	6.6 (5.6)	5.9 (10.1)	-0.7 (9.6)	6.9 (5.7)	1.6 (3.0)	-5.3 (6.2)	<0.001
	Median (Q1,Q3)	5 (3, 9)	2.2 (0.9, 6.5)	-2 (-4, 1)	5 (3, 9)	1 (0, 2)	-4 (-7, -2)	
ODI (N/hr)	N	208	208	208	199	199	199	
	Mean (SD)	8.2 (7.2)	7.2 (10.7)	-1.0 (9.9)	8.6 (7.6)	3.8 (4.1)	-4.8 (7.9)	<0.001
	Median (Q1,Q3)	6 (3, 11)	4 (2, 8)	-1.3 (-5, 1)	6 (3, 12)	3 (1, 5)	-3 (-8, -0)	
Time with End-Tidal CO_2 > 50 mm Hg (%TST)	N	146	146	146	148	148	148	
	Mean (SD)	9.0 (19.1)	9.5 (18.5)	0.5 (24.5)	12.0 (19.9)	7.3 (14.6)	-4.7 (20.8)	0.04
	Median (Q1,Q3)	1 (0, 5)	1 (0, 9)	0 (-1, 3)	2 (0, 14)	1 (0, 6)	-0 (-8, 1)	
Arousal index (N/hr)	N	208	208	208	199	199	199	
	Mean (SD)	8.4 (3.2)	8.6 (4.8)	0.2 (4.7)	8.6 (3.2)	7.2 (3.1)	-1.4 (3.9)	<0.001

		WWSC			eAT			
Outcome	**Parameter**	**Baseline**	**Month 7**	**Change from baseline to month 7**	**Baseline**	**Month 7**	**Change from baseline to month 7**	**p value**
	Median (Q1,Q3)	8 (6, 10)	8 (6, 10)	-0 (-3, 2)	8.0 (6.3, 10.3)	7 (5, 9)	-1.2 (-3.3, 0.4)	
Time in stage 1 (% TST)	N	208	208	208	199	199	199	
	Mean (SD)	8.7 (4.2)	8.4 (4.7)	-0.4 (4.7)	8.7 (4.2)	6.9 (3.2)	-1.8 (3.8)	<0.001
	Median (Q1,Q3)	8 (6, 11)	7 (5, 10)	-1 (-3, 2)	8 (6, 11)	6 (5, 9)	-1 (-4, 1)	
Time in stage 2 (% TST)	N	208	208	208	199	199	199	
	Mean (SD)	41.5 (7.9)	42.8 (7.0)	1.4 (7.8)	41.2 (7.5)	44.6 (7.3)	3.4 (8.2)	0.01
	Median (Q1,Q3)	42 (36, 47)	44 (38, 48)	1 (-4, 7)	41 (37, 46)	45 (39, 50)	3 (-2, 8)	
Time in stage N3 (% TST)	N	208	208	208	199	199	199	
	Mean (SD)	31.7 (7.6)	30.9 (6.7)	-0.8 (7.7)	31.4 (7.3)	30.0 (7.0)	-1.5 (8.2)	0.40
	Median (Q1,Q3)	32 (27, 35)	30 (26, 35)	-1 (-6, 5)	31 (26, 36)	29 (25, 34)	-1 (-6, 4)	
Time in stage REM (% TST)	N	208	208	208	199	199	199	
	Mean (SD)	18.1 (4.4)	17.9 (4.2)	-0.2 (4.9)	18.7 (4.2)	18.6 (4.0)	-0.1 (4.9)	0.76
	Median (Q1,Q3)	18.2 (16, 21)	18 (15, 21)	-0 (-3, 3)	19 (16, 22)	19 (16, 21)	-1 (-3, 3)	

All p values adjusted for the stratification factors of age (5-7 versus 8-9 years old), race (African American versus other), weight status (overweight/obese versus non-overweight) and study site. WWSC, Watchful Waiting With Supportive Care; eAT, early adenotonsillectomy group, NEPSY A/E, attention/executive function measured by the Developmental Neuropsychological

Assessment; Conners' CGI, Conner's Parent Rating Scale-Revised: Long version Global Index; BRIEF GEC, Behavior Rating Inventory of Executive Function Global Executive Composite T score; SRBD, Sleep Related Breathing Disorder Scale; PedsQL, Pediatric Quality of Life Inventory; AHI, apnea hypopnea index; ODI, oxygen desaturation index; TST, total sleep time; REM, rapid eye movement

Table S3: Adverse Events and Serious Adverse Events

Type of event (N)	**WWSC**	**eAT**
Tonsillar hemorrhage	1	2
Postoperative pain	0	3
Asthma	18	3
Lower respiratory tract	6	8
Upper respiratory tract (including ears)	90	67
Cough	11	15
Gastrointestinal tract	12	11
Dehydration	1	3
ADHD	4	3
Other infections (not included in other listed categories)	40	17
Hypersomnolence	2	1
Sleep apnea symptom exacerbation	6	0
Other	34	27
TOTAL	**225**	**160**
Serious Adverse Events		
Type of event (N)	**WWSC**	**eAT**
Tonsillar hemorrhage	1	3
Postoperative pain	1	1
Asthma	3	0
Lower respiratory tract	0	1
Upper respiratory tract	1	0

Vomiting/dehydration	1	1
Hypersomnolence	1	0
Hypertension	1	0
TOTAL	**9**	**6**

The number of adverse events and serious adverse events for the early adenotonsillectomy (eAT) and Watchful Waiting with Supportive Care Group (WWSC) groups are shown.

REFERENCES

1. Korkman N, Kirk U, Kemp S. NEPSY: A developmental neuropsychological assessment manual. New York: The Psychological Corporation; 1998.

2. Brodsky L. Modern assessment of tonsils and adenoids. Pediatr Clin North Am 1989;36(6):1551-1569.

3. Elliott CD. Differential Abilities Scale. San Antonio: Harcourt Assessment, Inc.; 1990.

4. Marcus CL, Omlin KJ, Basinki DJ et al. Normal polysomnographic values for children and adolescents. Am Rev Respir Dis 1992;146(5 Pt 1):1235-1239.

5. Witmans MB, Keens TG, Davidson Ward SL, Marcus CL. Obstructive hypopneas in children and adolescents: normal values. Am J Respir Crit Care Med 2003;168(12):1540.

6. Uliel S, Tauman R, Greenfeld M, Sivan Y. Normal polysomnographic respiratory values in children and adolescents. Chest 2004;125(3):872-878.

7. Traeger N, Schultz B, Pollock AN, Mason T, Marcus CL, Arens R. Polysomnographic values in children 2-9 years old: additional data and review of the literature. Pediatr Pulmonol 2005;40(1):22-30.

8. Montgomery-Downs HE, O'Brien LM, Gulliver TE, Gozal D. Polysomnographic characteristics in normal preschool and early school-aged children. Pediatrics 2006;117(3):741-753.

9. Kuczmarski RJ, Ogden CL, Grummer-Strawn LM et al. CDC growth charts: United States. Adv Data 2000;(314):1-27.

10. Himes JH, Dietz WH. Guidelines for overweight in adolescent preventive services: recommendations from an expert committee. The Expert Committee on Clinical Guidelines for Overweight in Adolescent Preventive Services. Am J Clin Nutr 1994;59(2):307-316.

11. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification. Iber C, editor. 2007. American Academy of Sleep Medicine.

12. Conners CK. Conners Rating Scales – Revised Technical Manual. North Tonawanda, NY: Multi-Health Systems; 1997.

13. Gioia GA, Isquith PK, Guy PK, Kenworthy L. Behavior Rating Inventory of Executive Function (BRIEF). Odessa, FL: Psychological Assessment Resources; 2000.

14. Chervin RD, Hedger K, Dillon JE, Pituch KJ. Pediatric sleep questionnaire (PSQ): validity and reliability of scales for sleep-disordered breathing, snoring, sleepiness, and behavioral problems. Sleep Med 2000;1(1):21-32.

15. Melendres MC, Lutz JM, Rubin ED, Marcus CL. Daytime sleepiness and hyperactivity in children with suspected sleep-disordered breathing. Pediatrics 2004;114(3):768-775.

16. Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. Med Care 2001;39(8):800-812.

17. Franco RA, Jr., Rosenfeld RM, Rao M. First place--resident clinical science award 1999. Quality of life for children with obstructive sleep apnea. Otolaryngol Head Neck Surg 2000;123(1 Pt 1):9-16.

18. Gottlieb DJ, Chase C, Vezina RM et al. Sleep-disordered breathing symptoms are associated with poorer cognitive function in 5-year-old children. J Pediatr 2004;145(4):458-464.

19. Krishna P, Lee D. Post-tonsillectomy bleeding: a meta-analysis. Laryngoscope 2001;111(8):1358-1361.

20. Tauman R, Gulliver TE, Krishna J et al. Persistence of obstructive sleep apnea syndrome in children after adenotonsillectomy. J Pediatr 2006;149(6):803-808.

21. Bhattacharjee R, Kheirandish-Gozal L, Spruyt K et al. Adenotonsillectomy outcomes in treatment of obstructive sleep apnea in children: a multicenter retrospective study. Am J Respir Crit Care Med 2010;182(5):676-683.

22. van Staaji BK, van den Akker EH, Rovers MM, Hordijk GJ, Hoes AW, Schilder AG. Effectiveness of adenotonsillectomy in children with mild symptoms of throat infections or adenotonsillar hypertrophy: open, randomised controlled trial. Clin Otolaryngol 2005;30(1):60-63.